Dist: A

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	2/9/95	Final 5/15/91 - 9/30/94

4. TITLE AND SUBTITLE
Pattern-Analysis Based Models of Masking by Spatially Separated Sounds

5. FUNDING NUMBERS
G
AFOSR-91-0289
61102F
2313-CS

6. AUTHOR(S)
Robert H. Gilkey

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Wright State University
Dayton, OH 45435

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR-95-0104

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Lt Col Collins
AFOSR/NL
110 Duncan Ave, Suite B115
Bolling AFB, DC 20332-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT *(Maximum 200 words)*

Work is presented on masked detection, sound localization, neural networks, and the sense of presence. Both free-field and headphone-based studies of masking indicate that traditional models of binaural interaction may be inadequate to explain the reductions in masking that often occur with dichotic, as opposed to diotic, stimulation. The accuracy of localization judgments in the presence of a masker is determined by both the signal-to-noise ratio and the spatial location of the masker. Localization judgments for speech stimuli are in general less accurate than those for some nonspeech stimuli. Neural-network models of sound localization can achieve performance similar to human observers based on monaural information alone or based on interaural information alone. The reports of suddenly-deafened adults suggest that realistic auditory stimulation may be critical for determining the sense of presence in virtual environments.

14. SUBJECT TERMS		15. NUMBER OF PAGES
Psychoacoustics	Virtual Reality	25
Spatial Hearing	Neural Networks	16. PRICE CODE
Auditory Masking		

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
U	U	U	U

Pattern-Analysis Based Models of Masking by Spatially Separated Sound Sources
AFOSR 91-0289
Final Technical Progress Report
May 15, 1991 to September 30, 1994

I. SUMMARY OF MAJOR FINDINGS

Research is described in three areas: masked detection, sound localization, and neural network-based models of sound localization. Our work on masked detection indicates that substantial reductions in masking of 8 to 18 dB can be realized when the signal is spatially separated from the masker in the free field. In general, these reductions in masking appear to be mediated by high-frequency monaural information, rather than by low-frequency binaural information. Headphone-based studies of reproducible noise masking question traditional models of binaural masking, by showing unexpected relations between responses under monaural and binaural conditions. A new response technique has been developed to support our work on sound localization. Sound localization performance degrades nearly monotonically with reductions in signal-to-noise ratio. In addition, localization judgments for the signal are systematically biased toward the direction of the masker. Localization judgments for speech stimuli are less accurate than localization judgments for click-train stimuli. Neural network models of sound localization can produce responses comparable to those of human observers. This human-like performance can be achieved based on either monaural or interaural stimulus cues. Library research suggests that auditory cues may be critical to the sense of presence in virtual environments. Our efforts in laboratory development and in planning and presenting the Conference on Binaural and Spatial Hearing are also briefly described.

II. RESEARCH OBJECTIVES

The long-term goal of this program of research is to specify the mechanisms that underlie the spatial hearing abilities of humans. Several projects are concerned with spatial hearing performance in the presence of noise. The motivation for this research is two fold. First, we are answering a series of basic science questions concerned with the mechanisms that allow us to "hear out" and process one particular stimulus, in the presence of other interfering stimuli. Second, the results relate to a series of applied questions, concerning the effectiveness of three-dimensional virtual auditory displays, when those displays are complex (e.g., containing many stimuli) or when they are used in a noisy environment (e.g., a cockpit). Other projects are comparing the accuracy of localization judgments for speech and nonspeech stimuli and evaluating the role of audition for creating a sense of presence in virtual environments. Another project is developing a model of spatial hearing. A number of potential monaural and binaural cues have been suggested as a potential basis for sound localization. We view the observer in a sound localization task as attempting to associate the pattern of acoustic cues received on a particular trial with a particular source direction. We are using neural network models to perform this pattern recognition task, and are attempting to determine which cues are necessary to achieve human-like performance.

III. STATUS OF THE RESEARCH

Much of the work described here is being conducted in the Auditory Localization Facility of the Armstrong Laboratory at Wright-Patterson Air Force Base. This facility contains a 14-foot

diameter geodesic sphere, with 277 speakers mounted on its surface. This is a unique facility that allows the experimenter considerable control over the spatial distribution of sound sources when conducting sound localization or free-field masking research. Additional studies are being performed in the Signal Detection Laboratory of the Department of Psychology at Wright State University. This is a more traditional psychoacoustic facility, where subjects listen to sounds presented over headphones in individual sound-attenuating booths. Some of the work described here receives additional support from Armstrong Laboratory, from a grant from the National Institutes of Health, and through cost-sharing funds from Wright State University.

A. Detection in Noise.

1. Free-field Masking. Our work on free-field masking replicates previous work that has shown a substantial increase in detectability when the signal and masker are spatially separated [e.g., K. Saberi, L. Dostal, T. Sadralodabai, V. Bull, and D.R. Perrott, J. Acoust. Soc. Am. 90, 1355-1370 (1991)]. However, in our work the stimulus frequency was systematically manipulated. In addition, free-field and headphone presentation of stimuli were compared, allowing us to assess the relative influence of monaural and binaural sources of information. In these studies, the detectability of a brief click-train signal in the presence of a white Gaussian noise masker was measured as a function of the spatial separation between the signal and the masker. Both the signal and the masker were band-limited to lie within low- (below 1.4 KHz), mid- (1.2 to 6.8 KHz), or high- (above 3.5 KHz) frequency ranges. Interaural time differences, interaural level differences, and spectral modulations introduced by the pinnae might be expected to be the most effective cues in the low-, mid-, and high-frequency regions, respectively. Three masker locations were examined: directly in front of the subject, directly to the left of the subject, and directly above the subject.

Our investigations in the free-field showed that when the signal was separated from the masker in azimuth within the horizontal plane, the detectability of the signal could be increased by as much as 18 dB (see Figure 1). Increases in detectability of as much as 8 dB were observed for separations in elevation within the median plane (see Figure 2). In all cases, the increases in detectability observed for the high-frequency signal were as great, or greater, than those observed for the low-frequency signal. Traditional models of binaural masking, based on interaural differences, did not predict the increases in detectability observed with vertical separations within the median plane, where interaural differences are relatively small. Moreover, these models seem inadequate to explain the effects of stimulus frequency; that is, the increase in the magnitude of the interaural level difference with increasing frequency is not great enough to predict the observed improvement in performance.

In a further attempt to relate these results to traditional headphone-based binaural masking results, continuous and gated maskers were compared. Based on the headphone results of McFadden [J. Acoust. Soc. Am. 40, 1414-1419 (1966)], who found that the binaural release from masking was greater with a continuous masker than with a gated masker, it might be expected that the effects of spatial separation would be greater with a continuous masker than with a gated masker. On average, we found that the signal was about 2-3 dB more detectable in the presence of a continuous masker. However, in conflict with the headphone results, these effects were not systematically related to spatial separation.

The effects of spatial separations were compared for "real" and "virtual" sounds, in order to determine the relative importance of monaural and binaural cues for detection. The virtual sounds were generated by passing the source waveforms through head-related transfer functions, which reproduced the direction-specific filtering of the head and pinnae that would be present in a real sound field. Because the stimuli were presented through headphones, monaural and binaural presentations could be compared by merely turning off one channel. Although there was some

evidence suggesting a small role for interaural cues at low frequencies, in most cases the best monaural performance was as good as binaural performance, suggesting that the increases in detectability observed in the free field could have been mediated by monaural changes in the effective signal-to-noise ratio, rather than by changes in interaural information.

Overall, the results of these studies indicate that reductions in masking on the order of 8 to 18 dB can be observed in free-field masking situations when the signal and the masker are spatially separated. The pattern of results from these experiments emphasizes the importance of high-frequency monaural information. These results have important implications for display designers, indicating the nature and magnitude of the changes in detectability that can be expected when sounds are spatially separated. Portions of this work were presented at the Boston University Binaural Conference, December 1991; the fall meeting of the Acoustical Society of America in 1992; the meeting of the Human Factors Society, October 1992; the AFOSR Review: Research on Hearing, June 1993; the Conference on Binaural and Spatial Hearing, September 1993; and the spring meeting of the Acoustical Society of America, June 1994. A proceedings paper describing some of this work has been published: Good and Gilkey (1992). Two papers describing this work are in press: Gilkey and Good (1995) and Good, Gilkey, and Ball (1995). A third paper is in revision: Good, Gilkey, and Ball (1995).

2. Reproducible Noise Masking. We have also been using headphone-based masking experiments to test the predictions of traditional models of binaural interaction. The large masking level difference (MLD) observed between monaural and binaural tone-in-noise masking tasks has been used to suggest that quite different processing is employed under these two conditions (e.g., energy detection vs. interaural time processing). However, when Gilkey, Robinson, and Hanna [J. Acoust. Soc. Am. 78, 1207-1219 (1985)] examined the trial-by-trial responses of subjects, they found that the responses under the N0S0 (both noise and signal presented diotically) and N0Sπ (noise diotic, but signal presented 180° out of phase interaurally) conditions were highly correlated. That is, although the signal level under the N0Sπ condition is 10 to 15 dB lower than under the N0S0 condition (because of the MLD), individual reproducible noise-alone or signal-plus-noise waveforms that were likely to elicit a positive response (i.e., a report of signal present) under one condition were also likely to elicit a positive response under the other condition. Gilkey et al. used wideband reproducible noise samples as maskers. When Isabelle and Colburn [J. Acoust. Soc. Am. 89, 352-359 (1991)] examined the responses of subjects to narrowband reproducible noise samples, they found correlations that were much weaker and often negative. They attributed the differences between their data and those of Gilkey et al. to the differences in the bandwidth of the masker. However, Gilkey [Paper presented at the Midwinter Meeting of the Association for Research in Otolaryngology, February, 1990] directly compared narrowband and wideband results for the same subjects and found highly significant correlations between N0S0 and N0Sπ conditions with both wideband and narrowband maskers. The correlation between N0S0 and N0Sπ responses has significant implications for models of both monaural and binaural performance. Lateralization-based models of binaural masking are unable to predict these data. On the other hand, Gilkey et al. showed that for a model such as the Equalization-Cancellation (EC) model [N.I. Durlach, "Binaural signal detection: Equalization and Cancellation Theory," in Foundations of Modern Auditory Theory II, edited by J.V. Tobias (Academic Press, New York), 371-462 (1972)] the effective maskers under the N0S0 and N0Sπ conditions are highly correlated. Thus, the observed correlation of responses between these conditions is not necessarily unexpected.

In order to evaluate more fully the predictions of the EC model, we have measured the responses of subjects to individual reproducible waveforms under conditions where the effective masker at the output of the EC device should be quite different from the masker under the N0S0 condition. Under the NuSπ condition (independent noises to the two ears, signal 180° out of

phase interaurally), the EC device should subtract the stimuli arriving from the two ears, such that the effective masker at the output of the EC device is the difference between the two monaural maskers. Thus, the EC model predicts that N0S0 responses to either of the two "monaural" maskers should be only partially correlated with the NuSπ responses. This is exactly what we observed in the data of our subjects. However, when the responses to the two monaural waveforms were averaged and then compared to the NuSπ responses, a very high correlation was observed. That is, the responses in the NuSπ condition could be predicted without "interaural" processing. This result was not anticipated.

Because we were surprised by this result, we next examined a condition where the actual masker under the N0S0 condition was the difference between the two maskers presented under the NuSπ condition. That is, the masker under the N0S0 condition was the predicted effective masker under the NuSπ condition based on the EC model. Rather than the strong correlation we expected between these two conditions, only a weak relation was observed.

Overall, the pattern of results suggested two possibilities, either: 1) the NuSπ condition is not a true binaural condition, as suggested by Durlach, Gabriel, Colburn, and Trahiotis [J. Acoust. Soc. Am. 79, 1548-1557 (1986)], or 2) the EC model is an inadequate model of binaural hearing. Parts of this work were presented at the Boston University Binaural Conference, December, 1991; the fall meeting of the Acoustical Society of America, October, 1992; and at the Midwinter Meeting of the Association for Research in Otolaryngology, February, 1993.

B. Localization in Noise.

In many situations, the observer must be able to determine the direction of the sound source, once it has been detected. Presumably, because of the increased complexity of the localization task relative to the detection task, a more complete representation of the signal information is needed for accurate localization. Much of our work on sound localization is represented by the thesis research of Michael D. Good. He is attempting to determine how localization performance is influenced by noise and how these performance changes can be related to our data on detection in noise.

1. Localization Pointing Technique. In designing these experiments, we realized that currently available response techniques allow responses to be collected at only very slow rates (2-4 responses per minute). It was clear that if we were to collect data at these slow rates, even relatively simple experiments would take months or years to complete. Therefore, we needed to develop a technique that would allow subjects to record accurately the perceived location of a sound, at speeds that were much more rapid than was possible with current techniques.

After considering a number of alternatives, we decided that a pointing technique would be the most effective procedure. It was known that subjects could accurately indicate the location of a sound by verbally reporting spherical coordinates [F.L. Wightman and D. Kistler, J. Acoust. Soc. Am. 85, 868-878 (1989)]. Our own pilot studies indicated that when presented with spherical coordinates, subjects could point to the corresponding location on a small plastic hemisphere to within a few degrees. We therefore developed a technique in which the subjects indicate the perceived location of a sound by pointing at an 8-inch spherical model of auditory space. The subjects point at the sphere using a magnetic stylus, whose XYZ coordinates are monitored with a Polhemus Fastrack "head tracker"; they then press a foot-switch to record their response. The results indicate that subjects are able to respond at rates of 16-19 responses per minute, considerably faster than with other techniques. Further, the accuracy of their responses is comparable to that which Wightman and Kistler observed with the verbal reporting technique.

Figures 3 and 4 compare the azimuth and elevation judgment centroids of our subjects to the judgment centroids of two of the subjects of Wightman and Kistler. These results were presented at the AFOSR Review: Research on Hearing, June 1993. A paper describing this technique has been published: Gilkey, Good, Ericson, Brinkman, and Stewart (1995).

2. **Effects of Signal-to-Noise Ratio.** In our first experiment on sound localization in noise, the subject's task was to localize a click-train signal that could originate from any of 239 directions surrounding the subject in azimuth and ranging from -45° to +90° in elevation. During each trial, a broadband Gaussian noise masker was presented from the speaker directly in front of the subject within the horizontal plane. Detection performance was measured in the quiet and at nine signal-to-noise ratios, ranging from -13 dB to +14 dB relative to the detection threshold for the signal when presented through the same speaker as the masker. The accuracy of the localization judgments decreased nearly monotonically as the signal-to-noise ratio was decreased. However, some aspects of the subjects' localization judgments were quite accurate at signal-to-noise ratios where other aspects were seriously degraded. That is, the accuracy of their judgments relative to the frontal plane (the Front/Back dimension) is disrupted even at relatively high signal-to-noise ratios; however, the accuracy of their judgments relative to either the median plane (the Left/Right dimension) or the horizontal plane (the Up/Down dimension) is not similarly disrupted, unless the signal-to-noise ratio is reduced considerably. These results are summarized in Figure 5. There are important implications of these results for the design of auditory displays. Information about the laterality of the signal, whether it is to the left or to the right of the user, is likely to be faithfully represented even in adverse environments. Elevation information, whether the signal is above or below the user, will be represented with similar fidelity. However, we can anticipate that users will have difficulty determining whether the signal is in front of them or behind them when the signal-to-noise ratio is unfavorable. Portions of this work were presented at the Conference on Binaural and Spatial Hearing, September 1993; Boston University Binaural Conference, December 1993; and the spring meeting of the Acoustical Society of America, June 1994. A chapter describing this work is in press: Good, Gilkey, and Ball (1995). A paper has been submitted: Good and Gilkey (1995a).

3. **Effects of Masker Location.** In a second experiment, the location of the masker was systematically varied. In different blocks of trials, the masker could be in front of the subject, behind the subject, directly to the subject's left, directly to the subject's right, or directly above the subject. At low signal-to-noise ratios, the subject's judgments of the direction of the signal were, in general, biased toward the direction of the masker. Note, however, that the location of the masker influences this pattern of results in a complex manner. Figures 6-8 give some indication of this complexity. For some combinations of masker location, signal location, and signal-to-noise ratio, responses appear to be biased away from the masker. Some masker locations appear to have a more general disruptive effect on localization performance (e.g., the masker location above the subject's head). Our examinations of the data from this experiment suggest that the pattern of results observed in the experiment described in Section III.B.2 is partially dependent on the location of the masker. Although performance in the Front/Back dimension is generally worse than in the other two dimensions, the decrease in performance as signal-to-noise ratio is lowered is most rapid when the masker is in front of the subject or behind the subject.

Interpretation of these data is complicated by the fact that all signals in this experiment were not presented at equal sensation levels. That is, when the signal was presented at a location close

to the masker it was less detectible than when the signal was presented from a location far from the masker. Thus, the detectability of the signal may have provided location information. Moreover, within any given block, some signals were likely to be well above detection threshold, whereas other signals were likely to be well below detection threshold; therefore, it is difficult to judge the relation between localization accuracy and the detectability of potential acoustic cues without analyzing the data on a location-by-location basis. However, even a location-specific analysis is difficult in this case, because detection data were not available for these same subjects or for most of the 239 spatial locations. Portions of this work were presented at the Conference on Binaural and Spatial Hearing, September 1993; Boston University Binaural Conference, December 1993; and the spring meeting of the Acoustical Society of America, June 1994. A chapter describing this work is in press: Good, Gilkey, and Ball (1995). A paper describing this work is in preparation: Good and Gilkey (1995b).

C. Localization of Speech Stimuli

Although most research on sound localization has been performed using nonspeech stimuli, in many auditory displays it will be necessary for users to accurately localize speech stimuli. Previous studies with speech stimuli have considered the accuracy of subjects' azimuth judgments, but have not systematically investigated the accuracy of elevation judgments. During the tenure of AFOSR-91-0289, we recorded a corpus of speech stimuli that were used during the fall of 1994 to measure the ability to localize speech. Subjects' accuracy with click-train stimuli was comparable to that with speech stimuli in the Left/Right dimension. However, judgments to click-train stimuli were consistently more accurate in the Front/Back dimension and typically more accurate in the Up/Down dimension. These results indicate that localization performance in applied settings, using speech stimuli, may be less accurate than would be expected based on the bulk of the previous literature. A paper describing this research has been submitted: Gilkey and Anderson (1995).

D. Neural Network Models of Sound Localization.

At least three sources of acoustic information are generally recognized as providing the foundation for sound localization: interaural time differences, interaural level differences, and direction-specific spectral modulations introduced by the acoustics of the torso, head, and pinnae. No model has been developed to describe how these disparate sources of information are combined into a single unified perception of the source location.

If the pattern of interaural time differences, interaural level differences, and spectral modulations is unique for each source direction, then the task of the observer in a localization experiment can be viewed as estimating the value of these cues and determining the location that corresponds to the estimated pattern; that is, within this view, sound localization is a pattern recognition task. Because neural networks have had great success in solving other pattern recognition problems, we have been using them to model sound localization.

Our models have been composed of a preprocessing section and a neural network section. In the preprocessing stage, the click signals were convolved with head-related transfer functions (filters that simulate the acoustic effect of the torso, head, and pinnae). The filtered clicks were then corrupted by internal noise; each point on the waveform was multiplied by a random amplitude gain (amplitude jitter) and subjected to a random delay (time jitter), in a manner similar to

that described by Durlach [J. Acoust. Soc. Am. 35, 1206-1218 (1963)]. A broadband cross-correlation was computed between the jittered waveforms in the left and right ears and the lag corresponding to the maximum in the cross-correlation function was one possible input to the network section of the model. In addition, Fast Fourier Transforms were computed from the waveforms in the left and right channels and the energy in each of 22 rectangular quarter-octave bands were determined. Logarithms of these quarter-octave spectra, or the difference between the log spectra in the left and right ears, were also possible inputs to the network stage. The network section was composed of 1 to 67 input units, followed by 0 to 50 hidden units and 30 output units.

In our initial investigations, the log-difference spectrum provided 22 inputs to the network and the interaural time delay corresponding to the maximum of a broadband cross-correlation provided the 23rd input. The sound source could originate from any of 144 directions, ranging in azimuth from -165° to +180° and in elevation from -36° to +54°. One hundred training vectors were generated for each of the 144 source locations (a total of 14,400 training vectors). A second set of 14,400 vectors was used as a test set. There were 23 input units, 50 hidden units, and 30 output units. A fully connected feed-forward network was trained, with back-propagation, to "turn on" 1 of 6 output units to indicate which of the 6 possible elevations had been presented, and 1 of 24 output units to indicate which of the 24 possible azimuths had been presented.

Figure 9 shows a comparison of the responses of a human subject and of the model in comparable listening conditions. The top two panels show the azimuth component of the judgment centroid plotted as a function of the target azimuth. As can be seen, both the network model and the human subject made very accurate azimuth judgments. The bottom two panels show the elevation component of the judgment centroids as a function of the actual elevation. The overall performance of the human and the model are similar, but the human systematically overestimates the elevation, while the model underestimates high elevations and overestimates low elevations. (This results, in part, from the response restrictions placed on the model; that is, it cannot respond with elevations greater than 54° or less than 36°; the human was not similarly constrained). The average angle of error for the human and the model are similar in magnitude and the number of front/back reversals observed for the model and the human are also comparable.

Wightman and Kistler [J. Acoust. Soc. Am. 91, 1648-1661 (1992)] demonstrated that for human observers low-frequency timing information plays a dominant role in determining their localization judgments. That is, when interaural time cues provide information about the source location that conflicts with information provided by interaural level differences or "spectral cues," subjects tend to judge the sound as coming from the location indicated by the interaural time difference, rather than from the location indicated by these other cues. Following Wightman and Kistler, we took the previously described network (trained on "normal" stimuli) and tested it on stimuli with phase spectra that had been modified to correspond to the phase spectra of a sound coming from 0° azimuth and 0° elevation, from -45° azimuth and 0° elevation, or from 90° azimuth and 0° elevation. The pattern of errors observed for the model was quite similar to that observed for Wightman and Kistler's human subjects. That is, in most cases, the model responded with the location indicated by the phase spectra, rather than the location indicated by the power spectra.

There has been some controversy over whether monaural or interaural cues form the basis for sound localization. Given that this model received only interaural inputs, there was no way for the model to recover the original monaural spectra. Thus, these results suggest that interaural information is sufficient to achieve localization performance comparable to that of humans.

A "pure" monaural model can also achieve performance similar to human performance. Separate networks were trained to localize based on the spectrum in the left ear and based on the

spectrum in the right ear. That is, each network had 22 input units with activation levels corresponding to the logarithm of the energy at the output of each 22 quarter-octave spectral bands from one ear. The outputs of these networks (i.e., the activation levels of the nodes corresponding to the 24 azimuths and 6 elevations) were used as inputs to a third, arbitrator, network. This hierarchical network performed as well or better than humans. In this case, binaural interaction, in the traditional sense, was not possible. That is, the information that was combined between the two ears was on weighted information about the models' location judgments, not detailed information about the stimulus waveforms or spectra. Therefore, the results of this experiment indicate that performance comparable to human performance can be achieved with a model that receives only monaural information.

It is important to note that, in this modeling effort, we use the neural network in a role similar to that of an "ideal detector"; thus, the implications of this work are not in terms of the structure of the neural network itself. Rather we are determining the viability of various acoustic cues for mediating human-like performance. Thus far, this work indicates that for the simple stimuli employed here there is sufficient information in either the monaural or interaural cues to produce localization performance comparable to that of humans. Portions of this work were presented at the fall meeting of the Acoustical Society of America in 1992; the Boston University Binaural Conference, December 1992; the AFOSR Review: Research on Hearing, June 1993; the Conference on Binaural and Spatial Hearing, September 1993; and the spring meeting of the Acoustical Society of America, June 1994. A proceedings paper describing some of this work has been published: Anderson, Janko, and Gilkey (1994). Two additional papers are in preparation: Janko, Anderson, and Gilkey (1995) and Anderson, Janko, and Gilkey (1995).

E. The Role of Auditory Stimulation in Achieving a Sense of Presence

Ramsdell ["The psychology of the hard-of-hearing and the deafened adult," in Hearing and Deafness, edited by S.R. Silverman and H. Davis (Holt, Rinehart, and Winston, New York), 499-510 (1978)] reports that adventitiously-deafened individuals feel a sense of unconnectedness with their surroundings, a sense that the world seems "dead." Such reports offer a compelling rationale for the argument that auditory cues are a crucial determinant of the sense of presence. Moreover, the crucial element of auditory stimulation for creating a sense of "presence" may be the auditory background, comprising the incidental sounds made by objects in the environment, rather than the communication and warning signals that typically capture our attention. Although designers of virtual environments have most often tried to maximize the sense of presence in the user by attempting to improve the fidelity of visual displays, we argue that background auditory stimulation may be useful or even critical for achieving a full sense of presence. A paper presenting this argument has been submitted: Gilkey and Weisenberger (1995).

F. Laboratory Development.

In February of 1991, the Signal Detection Laboratory was moved from the Central Institute for the Deaf (CID) in St. Louis to the Department of Psychology at Wright State University in Dayton. At this same time, Dr. Gilkey began his affiliation with the Armstong Laboratory at Wright-Patterson Air Force Base and initiated a program of research on free-field masking and sound localization. Both of these changes have required that considerable effort during this grant period be spent on laboratory development.

1. Armstrong Laboratory. A number of modifications to the Auditory Localization Facility at Wright-Patterson Air Force Base have been implemented to provide for more efficient data collection, higher fidelity sound production, and experiment-specific enhancements.

Previously, a single 80386/33 personal computer had been used for controlling the geodesic sphere, for controlling the localization cue synthesizers (used to present localized sound images through headphones), for hardware and software development on both systems, and for demonstrations for laboratory visitors. Hence, all of these activities were limited by the availability of a single personal computer. An 80486/33 personal computer was purchased to be dedicated exclusively to controlling experiments in the geodesic sphere.

Software drivers have been developed on the 80486 for sound generation, for sound processing and production, for controlling the sphere, programmable attenuators, and visual displays, and for recording subject responses. Because, as originally designed, the geodesic sphere did not provide a speaker directly in front of the subject, directly to the subject's left, directly to the subject's right, directly behind the subject, or directly above the subject, additional speakers were installed and the existing signal-switching hardware was modified to accommodate their presence.

The Auditory Localization Facility was initially designed to allow subjects to perform in localization experiments requiring them to turn and face the sound source. Therefore, the design allowed for a standing subject. For our detection experiments it is critical that the subject's head be stationary (i.e., a slight head movement can dramatically change the effective signal-to-noise ratio). After trying a number of alternatives it was determined that a stationary head could be best achieved for a seated subject using a bite-bar. This required the design of a chair, a bite-bar, and an extension to the stand.

A response box was built and interfaced to the computer. Because some of the experiments required continuous noise, it was necessary to have lights to mark the observation intervals. Thus, an LED display was constructed and installed on the surface of the sphere directly in front of the subject. An intercom and video camera were installed to increase the efficiency of data collection and to provide increased safety for the subjects.

Careful and consistent calibration of speakers is an important component of most localization and free-field masking research. We have made a number of enhancements and modifications to assure that an equivalent, spectrally-flat stimulus is produced by each speaker. A rotating microphone stand was developed which can, under computer control, turn the microphone to face each of the speakers. This allows the transfer function of all 239 speaker locations used in our localization research to be measured. Based on these transfer functions a digital filter is designed for each speaker such that when a spectrally flat stimulus is convolved with the filter and played through the speaker, the stimulus reaching the center of the sphere will also have a flat spectrum. After the filters have been designed, each one of them is tested and if the resulting spectrum varies by more than ± 0.5 dB within the 0.5 kHz to 10.1 kHz passband, the filter is modified and the procedure is repeated until this criterion is met. Because the transfer characteristics of the speakers are highly susceptible to temperature and humidity changes, a new set of transfer functions is measured every day and new filters are designed based on a weighted average of the transfer functions collected over the previous 5 days.

Computer software has been developed to perform an adaptive, cued, two-alternative, forced-choice task for our detection work. The signal level is varied adaptively to estimate the level that would produce 79.4% correct performance. The program allows selection of any two of the

speakers for presentation of the signal and masker or both signal and masker can be presented from the same speaker.

In support of our localization research, the new pointing response technique, described previously, was developed. Experiment control software was developed, which incorporates this technique, along with the ability to present sounds from random directions selected from a set of up to 239 speakers. In addition, a second source (i.e., a masker) can be presented from any of 5 fixed locations. Each stimulus is preconvolved with the appropriate transfer function such that when played through the selected speaker a spectrally flat stimulus will reach the center of the geodesic sphere.

2. Wright State University. Wright State remodeled the rooms that house the laboratory and provided four individual IAC sound attenuating booths. Computer hardware and software necessary for experimental control were installed. Necessary analog hardware that was not brought from CID has been purchased or built. The initial setup was functionality equivalent to that of the laboratory at CID. Data collection began in the fall quarter of 1991. More recently we have installed a 80386-based personal computer and "state of the art" Tucker-Davis Technologies audio generation hardware to replace our aging PDP11-based system. This system provides us with increased reliability, as well as enhanced hardware and software compatibility with the PC/Tucker-Davis systems on the base.

A DECstation 5000 that was brought from CID, and two DECsystem 3000 Model 400 computers purchased with AFOSR funds have been installed and connected to the campus Ethernet network. Two X-terminals and a large-screen Macintosh computer are also attached to the network. A third X-terminal borrowed from Ohio State University has been installed in the Armstrong Laboratory to allow student experimenters access to X-services while running experiments on the base. These systems provide data analysis, graphics, and modeling capabilities for the laboratory.

G. Conference on Binaural and Spatial Hearing

The Conference on Binaural and Spatial Hearing was held at the Hope Hotel and Conference Center at Wright-Patterson Air Force Base , Ohio, on September 9-12, 1993. This conference was sponsored by AFOSR Task 2313V3 and Armstrong Laboratories. Personnel on AFOSR 91-0289 were involved in every detail of conference planning and presentation, from contacting potential speakers to setting up audio-visual equipment. Speakers at the conference included Timothy Anderson, Leslie Bernstein, Jens Blauert, John Brugge, Thomas Buell, Mahlom Burkhard, Robert Butler, Rachel Clifton, Steven Colburn, Theodore Doll, Richard Duda, Nathaniel Durlach, Raymond Dye, Mark Ericson, Scott Foster, Robert Gilkey, Wesley Grantham, Ervin Hafter, William Hartman, Janet Koehnke, Armin Kohlrausch, Birger Kollmeier, Gregory Kramer, Shigeyuki Kuwada, Richard McKinley, Donald Mershon, John Middlebrooks, David Perrott, Edgar Shaw, Kourosh Saberi, Barbara Shinn-Cunningham, Richard Stern, Elizabeth Wenzel, Frederic Wightman, Tom Yin, William Yost, and Eric Young. More than sixty additional non-speaking attendees registered for the conference and provided many valuable insights. A book loosely based on the conference is currently in preparation and should be published in 1995: Gilkey and Anderson (1995).

IV. PUBLICATION ACTIVITY

Papers Published

Gilkey, R.H., Good, M.D., Ericson, M.A., Brinkman, J., & Stewart, J.M. (1995). A pointing technique for rapidly collecting localization responses in auditory research. Behavior Research Methods, Instrumentation, and Computers, 27, 1-11.

Anderson, T.R., Janko, J.A., & Gilkey, R.H. (1994). Modeling human sound localization with hierarchial neural networks. Proceedings of the IEEE International Conference on Neural Networks, VII, 4502-4507.

Good, M.D., & Gilkey, R.H. (1992). Masking between spatially separated sounds. Proceedings of the 36th Annual Meeting of the Human Factors Society, 36, 253-257.

Papers in press

Gilkey, R.H., & Good, M.D. (1995). Effects of frequency on free-field masking. Human Factors, in press.

Good, M.D., Gilkey, R.H., & Ball, J.M. (1995). The relation between detection in noise and localization in noise in the free field. In R.H. Gilkey & T.R. Anderson (Eds.), Binaural and Spatial Hearing. Hillsdale, NJ: Erlbaum, in press.

Papers submitted

Gilkey, R.H., & Anderson, T.R. (1995). The accuracy of absolute localization judgments for speech stimuli. Journal of Vestibular Research, submitted.

Good, M.D., & Gilkey, R.H. (1995a). Sound localization in noise: I. Effects of signal-to-noise ratio. Journal of the Acoustical Society of America, submitted.

Gilkey, R.H., & Weisenberger J.M. (1995). The sense of presence for the suddenly-deafened adult: Implications for virtual environments. Presence: Teleoperators and Virtual Environments, submitted.

Gilkey, R.H., Good, M.D., & Ball, J.M. (1995). A comparison of "free-field" masking for real and for virtual sounds. Journal of the Acoustical Society of America, in revision.

Papers in preparation

Good, M.D., & Gilkey, R.H. (1995b). Sound localization in noise: II. Effects of masker position. Journal of the Acoustical Society of America, in preparation.

Janko, J.A., Anderson, T.R., & Gilkey, R.H. (1995). Neural network models of monaural and binaural sound localization. In R.H. Gilkey & T.R. Anderson (Eds.), Binaural and Spatial Hearing. Hillsdale, NJ: Erlbaum, in preparation.

Gilkey, R.H., & Anderson, T.R. (Eds.). (1995). Binaural and Spatial Hearing. Hillsdale, NJ: Erlbaum, in preparation.

Anderson, T.R., Janko, J.A., & Gilkey, R.H. (1995). Neural network modeling of sound localization based on binaural cues. Journal of the Acoustical Society of America, in preparation.

Theses and Dissertations

Good, M.D. (1994). The Influence of Noise on Auditory Localization in the Free Field, unpublished master's thesis, Wright State University.

V. PARTICIPATING PROFESSIONALS

Robert H. Gilkey

De Anza College, Cupertino, CA

University of California, Berkeley, CA	B.A.	1976	Psychology
Indiana University, Bloomington, IN	Ph.D.	1981	Psychology

Dissertation title: "Molecular psychophysics and models of auditory signal detectability."

VI. INTERACTIONS

Conference presentations and invited talks

Anderson, T.R., Janko, J.A., & Gilkey, R.H. (1994). Modeling human sound localization with hierarchial neural networks. Presented at the IEEE International Conference on Neural Networks, June, Orlando, FL.

Ball, J.M., Gilkey, R.H., & Good, M.D. (1994). Binaural and monaural influences on "free-field" masking for real and virtual sound sources. Journal of the Acoustical Society of America, 95, 2897 (A).

Gilkey, R.H., Janko, J.A., & Anderson, T.R. (1994). Neural network models of sound localization based on monaural information and based on binaural information. Journal of the Acoustical Society of America, 95, 2898 (A).

Good, M.D., & Gilkey, R.H. (1994). Auditory localization in noise. I: The effects of signal-to-noise ratio. Journal of the Acoustical Society of America, 95, 2896 (A).

Good, M.D., & Gilkey, R.H. (1994). Auditory localization in noise. II: The effects of masker location. Journal of the Acoustical Society of America, 95, 2896 (A).

Gilkey, R.H., & Good, M.D. (1993). Localization in noise. Boston University Binaural Conference, Boston, MA, December.

Gilkey, R.H., & Good, M.D. (1993). Masking between spatially separated sound sources. Conference on Binaural and Spatial Hearing, Dayton, OH, September.

Anderson, T.R., Janko, J.A., & Gilkey, R.H. (1993). Using neural networks to model human sound localization. Conference on Binaural and Spatial Hearing, Dayton, OH, September.

Gilkey, R.H. (1993). Pattern-analysis based models of masking by spatially separated sound sources. AFOSR Review: Research in Hearing, Fairborn, OH, June.

Gilkey, R.H. (1993). Comparing predictions of the Equalization-Cancellation Model to NuSπ performance in a reproducible noise masking task. Association for Research in Otolaryngology Mid-Winter Meeting, St. Petersburg Beach, FL, February.

Gilkey, R.H. (1993). Auditory space perception and virtual environments. Ohio Consortium for Virtual Environment Research, Dayton, OH, January.

Gilkey, R.H., Janko, J.A., & Anderson, T.R. (1992). Using neural nets to model sound localization. Boston University Binaural Conference, Boston, MA, December.

Gilkey, R.H., & Good, M.D. (1992). Effects of frequency and masker duration on free-field masking. Journal of the Acoustical Society of America, 92, 2334(A).

Anderson, T.R., Janko, J.A., & Gilkey, R.H. (1992) An artificial neural network model of human sound localization. Journal of the Acoustical Society of America, 92, 2298(A).

McKinley, R.L., Ericson, M., Perrott, D., Brungart, D., Gilkey, R., & Wightman, F. (1992). Minimum audible angle for synthesized localization cues presented over headphones. Journal of the Acoustical Society of America, 92, 2297(A).

Gilkey, R.H. (1992). The correlation between responses under monaural and binaural conditions. Journal of the Acoustical Society of America, 92, 2298(A).

Good, M.D., & Gilkey, R.H. (1992) Masking between spatially separated sounds. The 36th Annual Meeting of Human Factors Society, Atlanta, GA, October.

Gilkey, R.H. (1991). Headphone and free-field studies of binaural masking. Boston University Binaural Conference, Boston, MA, December.

VII. FIGURE CAPTIONS

Figure 1. Threshold signal-to-noise ratio is plotted as a function of the azimuth of the signal within the horizontal plane. The masker was presented from directly in front of the subject, 0° azimuth, 0° elevation (left panel), or from directly to the left of the subject, -90° azimuth and 0° elevation, (right panel). Threshold estimates have been averaged across the three subjects. Negative values along the abscissa indicate positions to the left of the subject and positive values indicate positions to the right of the subject. A value of 0° indicates a speaker location directly in front of the subject. The position of the arrow shows the location of the masker.

Figure 2. Threshold signal-to-noise ratio is plotted as a function of the elevation of the signal within the median plane. The masker was presented from directly in front of the subject, 0° azimuth, 0° elevation (left panel), or from directly above the subject, 0° azimuth and 90° elevation (right panel). Threshold estimates have been averaged across the three subjects. Negative values along the abscissa indicate positions below the horizontal plane and positive values indicate positions above the horizontal plane. Elevations greater than 90° indicate locations in the rear hemisphere. The position of the arrow shows the location of the masker.

Figure 3. The azimuth coordinate of the judgment centroid for each target location is plotted as a function of the azimuth of the target. The top three panels show data for each of the 3 subjects in our experiment; they responded with the pointing technique. The bottom two panels show data for 2 of the subjects of Wightman and Kistler; they responded verbally. (The panel on the bottom-left shows data from one of their better subjects and the panel on the bottom-right shows data from one of their worst subjects.) The centroids in the top panels are based on 8 judgments at each speaker location. The centroids in the bottom panels are based on either 6 or 12 judgments at each speaker location. Front-back reversals have been resolved.

Figure 4. The elevation coordinate of the judgment centroid for each target location is plotted as a function of the target elevation. Note that the range of values on the axes has been reduced substantially relative to Figure 3. Other details are as in Figure 3.

Figure 5. The root-mean squared error, between the subject's judgment vector and the actual target vector, is plotted as a function of signal-to-noise ratio, for the left/right, up/down, and front/back dimensions. Data are shown for each of three subjects and for the average subject.

Figure 6. Scatter plots of the left/right coordinate of the subject's judgment vector as a function of the left/right coordinate of the target vector, for each of five masker locations. The size of symbols indicates the proportion of responses in each 5° wide target angle bin that fell within each 5° wide judgment angle bin. The upper panel shows the results for a masker in front of the subject. The middle panel shows the results for a masker above the subject. The lower panel shows the results for a masker behind the subject. The left panel shows the results for a masker to the left of the subject. The right panel shows the results for a masker to the right of the subject. Data for subject JY are shown.

Figure 7. The up/down coordinate of the subject's judgment vector is plotted as a function of the up/down component of the target vector, for each of five masker locations. Other details are as in Figure 6.

Figure 8. The front/back coordinate of the subject's judgment vector is plotted as a function of the front/back coordinate of the target vector, for each of five masker locations. Other details are as in Figure 6.

Figure 9. Performance of subject SDO, from the study of Wightman and Kistler (1989), and performance of a neural network, receiving only interaural cues, are compared. In the top two panels, the azimuth coordinate of the judgment centroid is plotted as a function of the target azimuth. In the bottom two panels, the elevation coordinate of the judgment centroid is plotted as a function of the target elevation. The panels on the left show the results for subject SDO. The panels on the right show the results for the neural network.

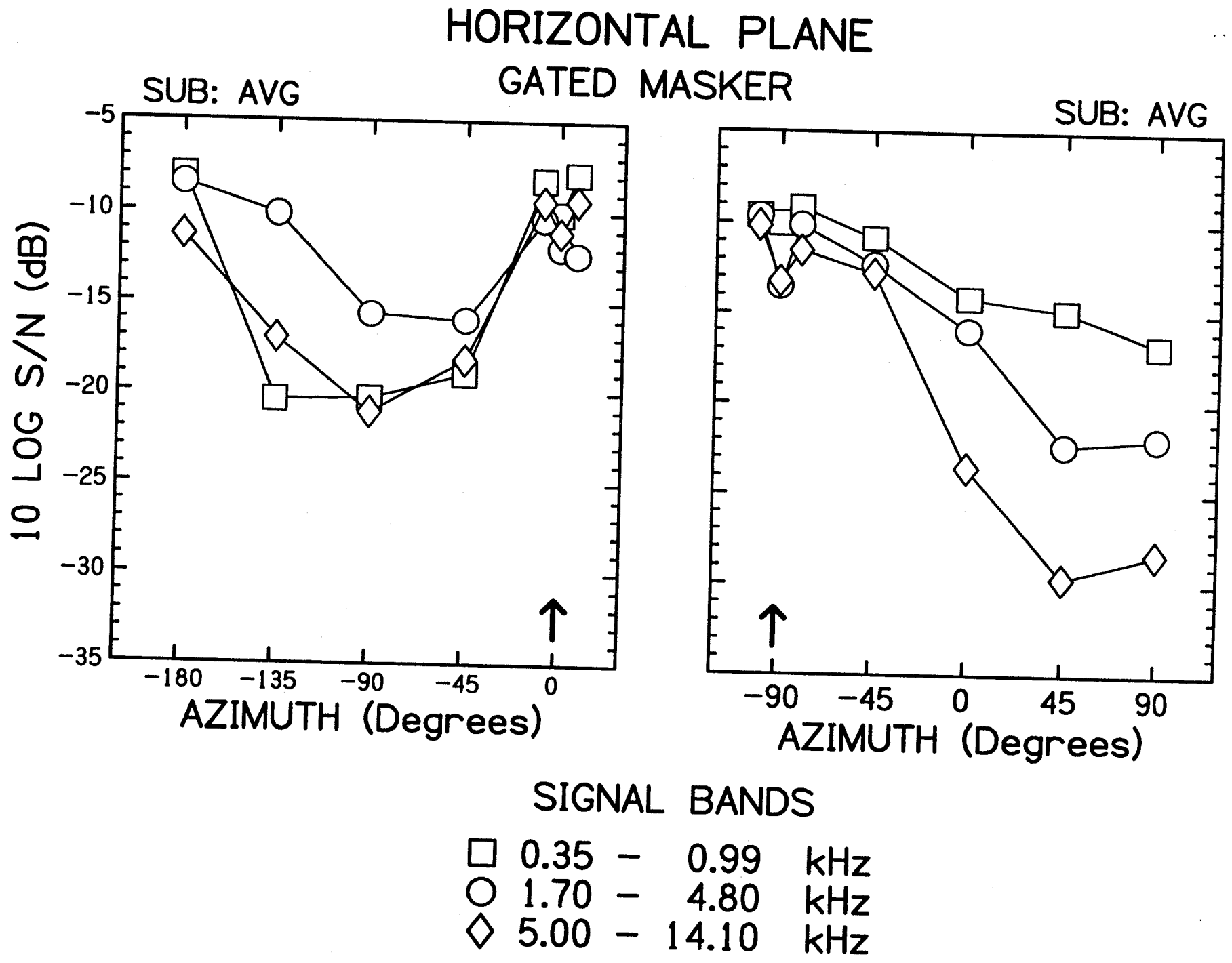

HORIZONTAL PLANE
GATED MASKER
SUB: AVG
SUB: AVG
10 LOG S/N (dB)
-5
-10
-15
-20
-25
-30
-35
-180
-135
-90
-45
0
AZIMUTH (Degrees)
-90
-45
0
45
90
AZIMUTH (Degrees)
SIGNAL BANDS
0.35 - 0.99 kHz
1.70 - 4.80 kHz
5.00 - 14.10 kHz

Figure 1

Figure 2

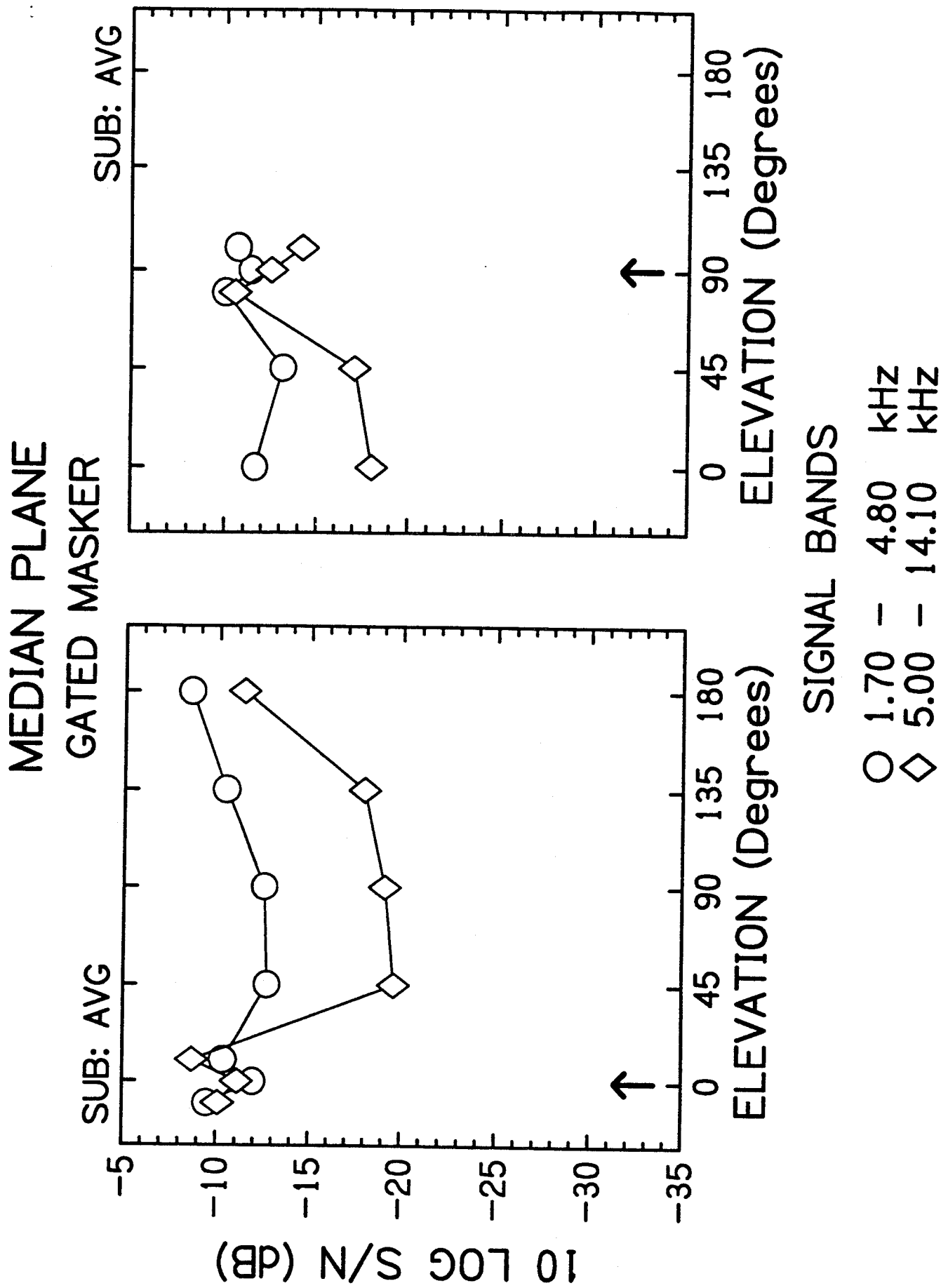

MEDIAN PLANE
GATED MASKER
SUB: AVG
SUB: AVG
10 LOG S/N (dB)
-5
-10
-15
-20
-25
-30
-35
0
45
90
135
180
ELEVATION (Degrees)
ELEVATION (Degrees)
SIGNAL BANDS
1.70 - 4.80 kHz
5.00 - 14.10 kHz

Figure 3

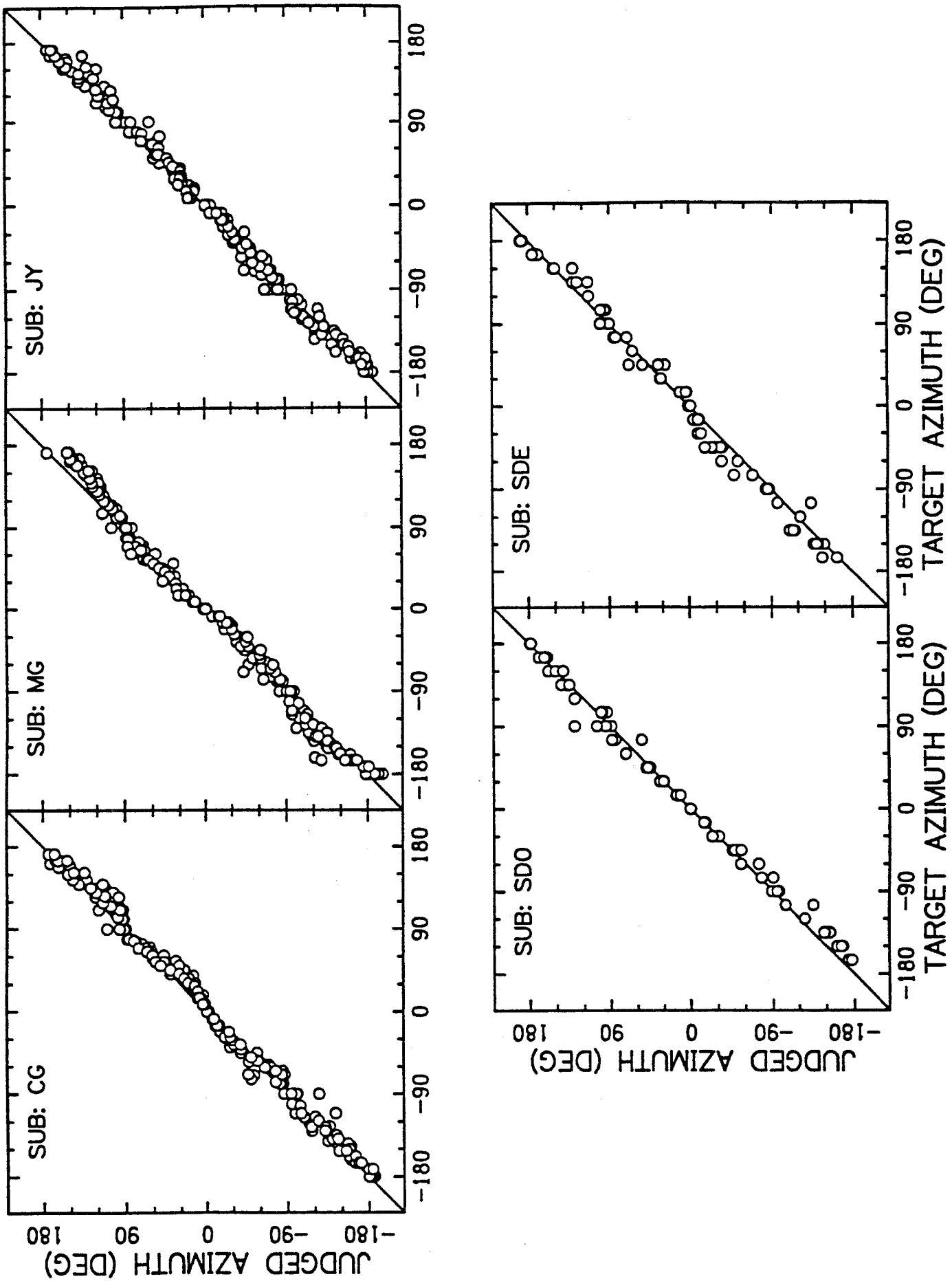
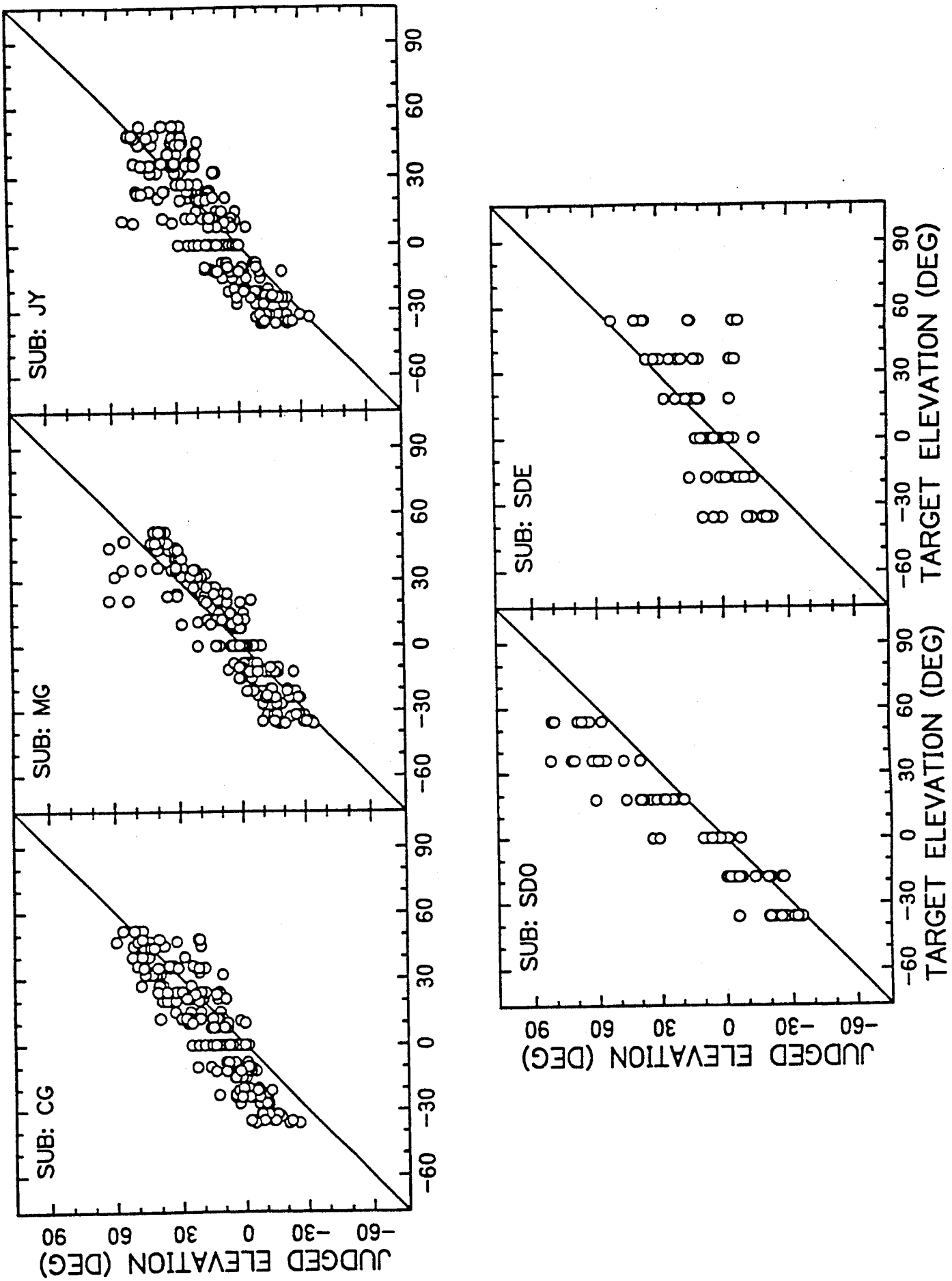
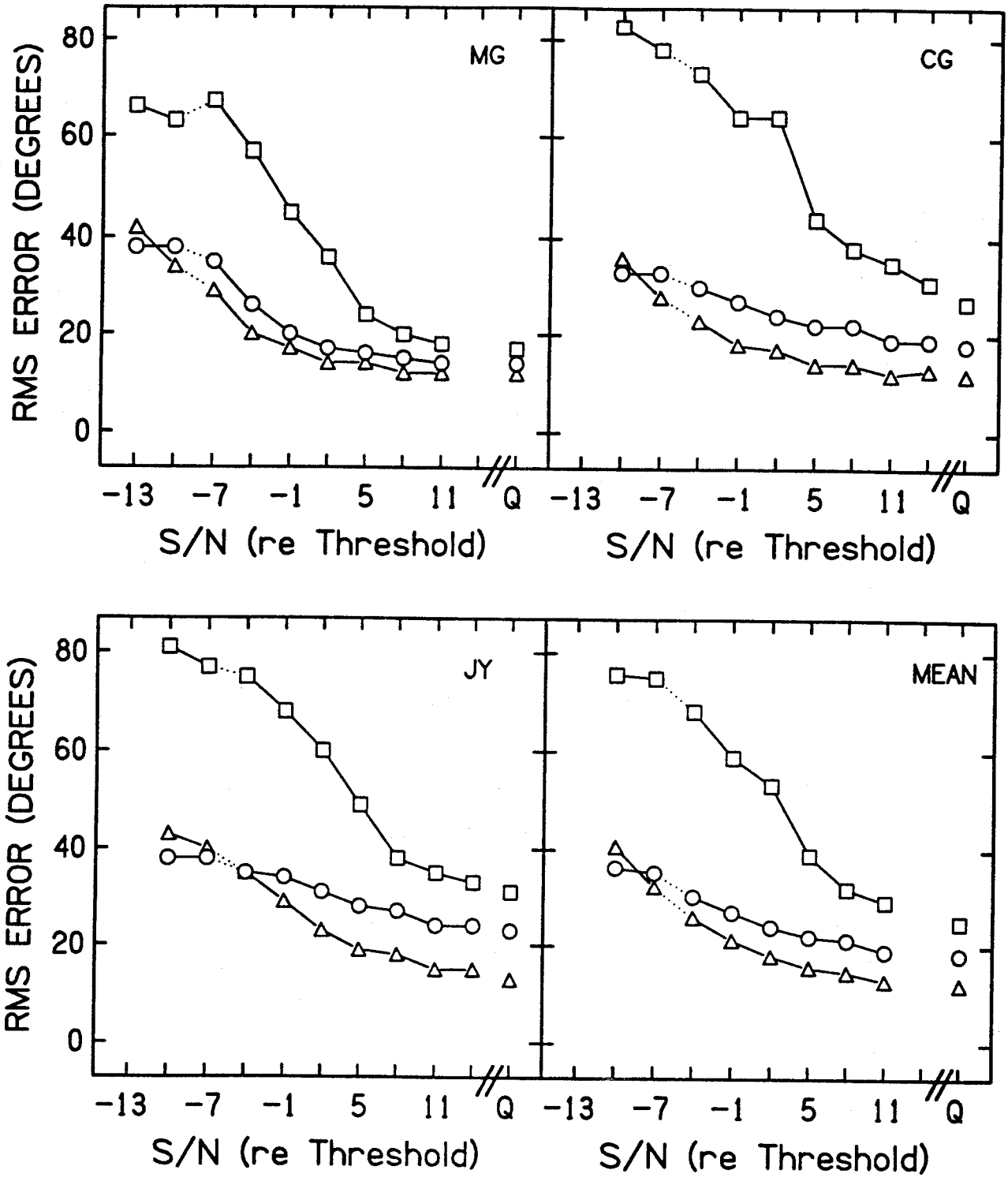
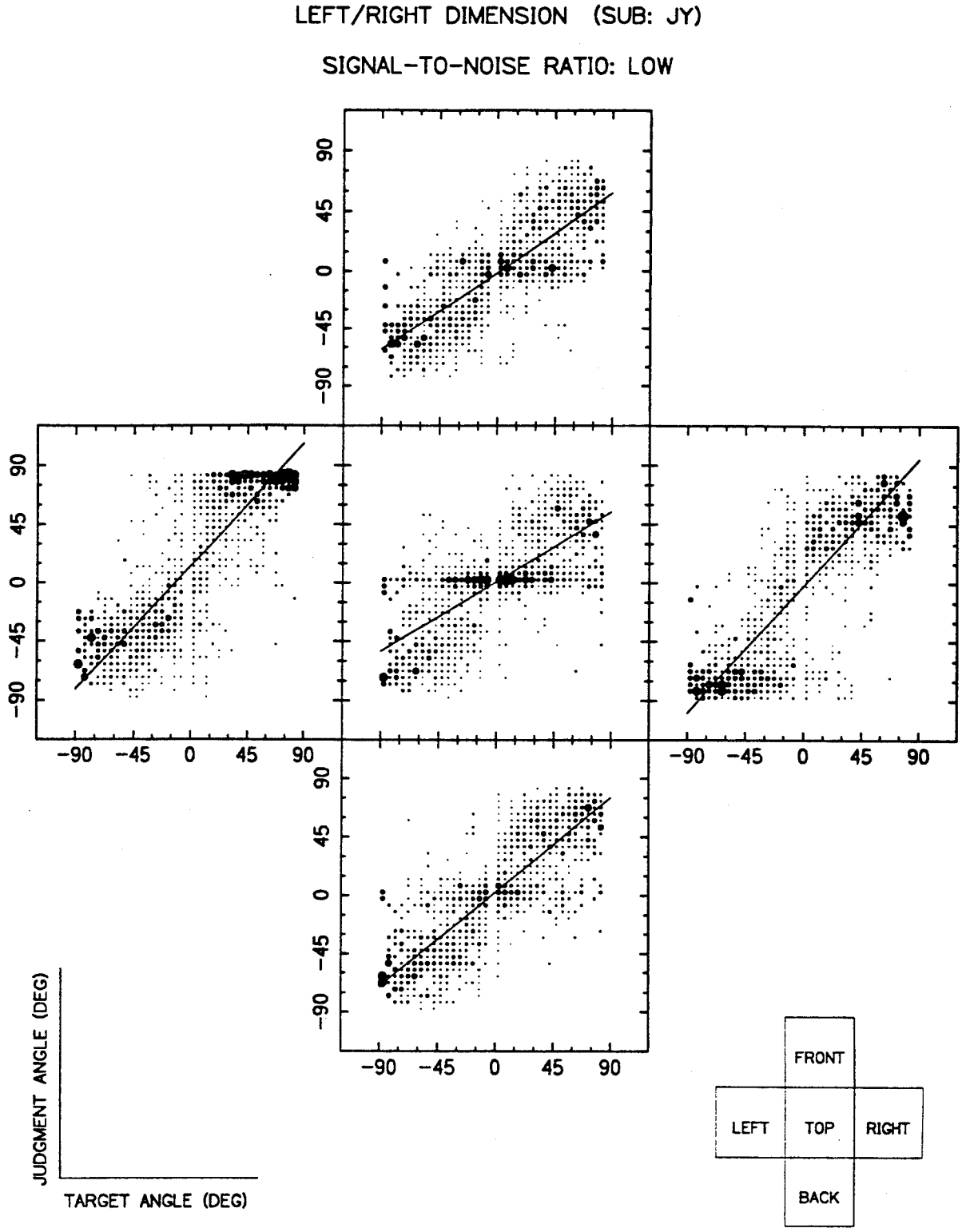
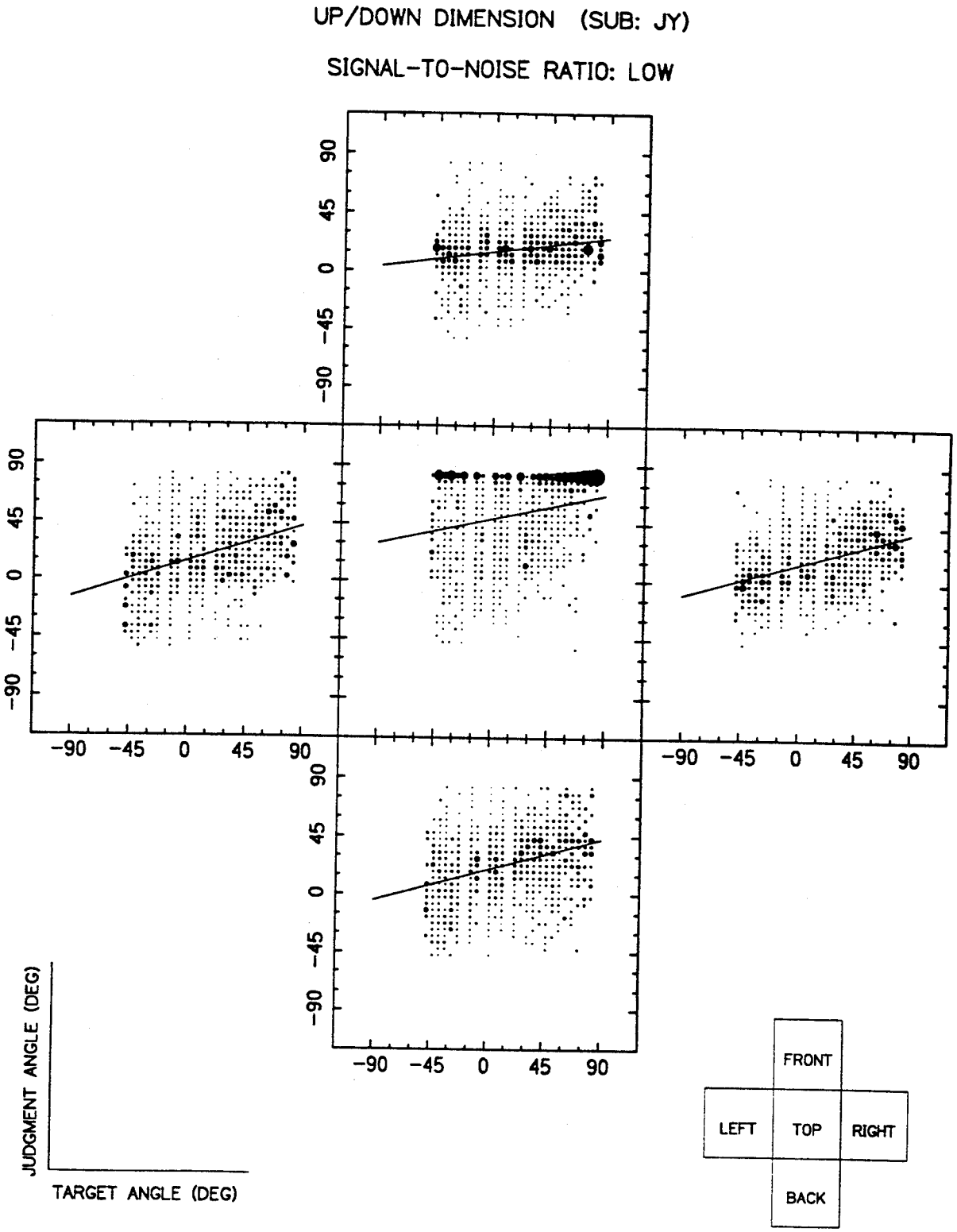

SUB: CG
SUB: MG
SUB: JY
SUB: SDO
SUB: SDE
JUDGED AZIMUTH (DEG)
TARGET AZIMUTH (DEG)
-180
-90
0
90
180

Figure 4


SUB: CG
SUB: MG
SUB: JY
SUB: SDO
SUB: SDE
JUDGED ELEVATION (DEG)
TARGET ELEVATION (DEG)
-60
-30
0
30
60
90


MG
CG
JY
MEAN
RMS ERROR (DEGREES)
80
60
40
20
0
-13
-7
-1
5
11
Q
S/N (re Threshold)
FRONT/BACK
UP/DOWN
LEFT/RIGHT


LEFT/RIGHT DIMENSION (SUB: JY)
SIGNAL-TO-NOISE RATIO: LOW
-90
-45
0
45
90
JUDGMENT ANGLE (DEG)
TARGET ANGLE (DEG)
FRONT
LEFT
TOP
RIGHT
BACK


UP/DOWN DIMENSION (SUB: JY)
SIGNAL-TO-NOISE RATIO: LOW
-90
-45
0
45
90
JUDGMENT ANGLE (DEG)
TARGET ANGLE (DEG)
FRONT
LEFT
TOP
RIGHT
BACK

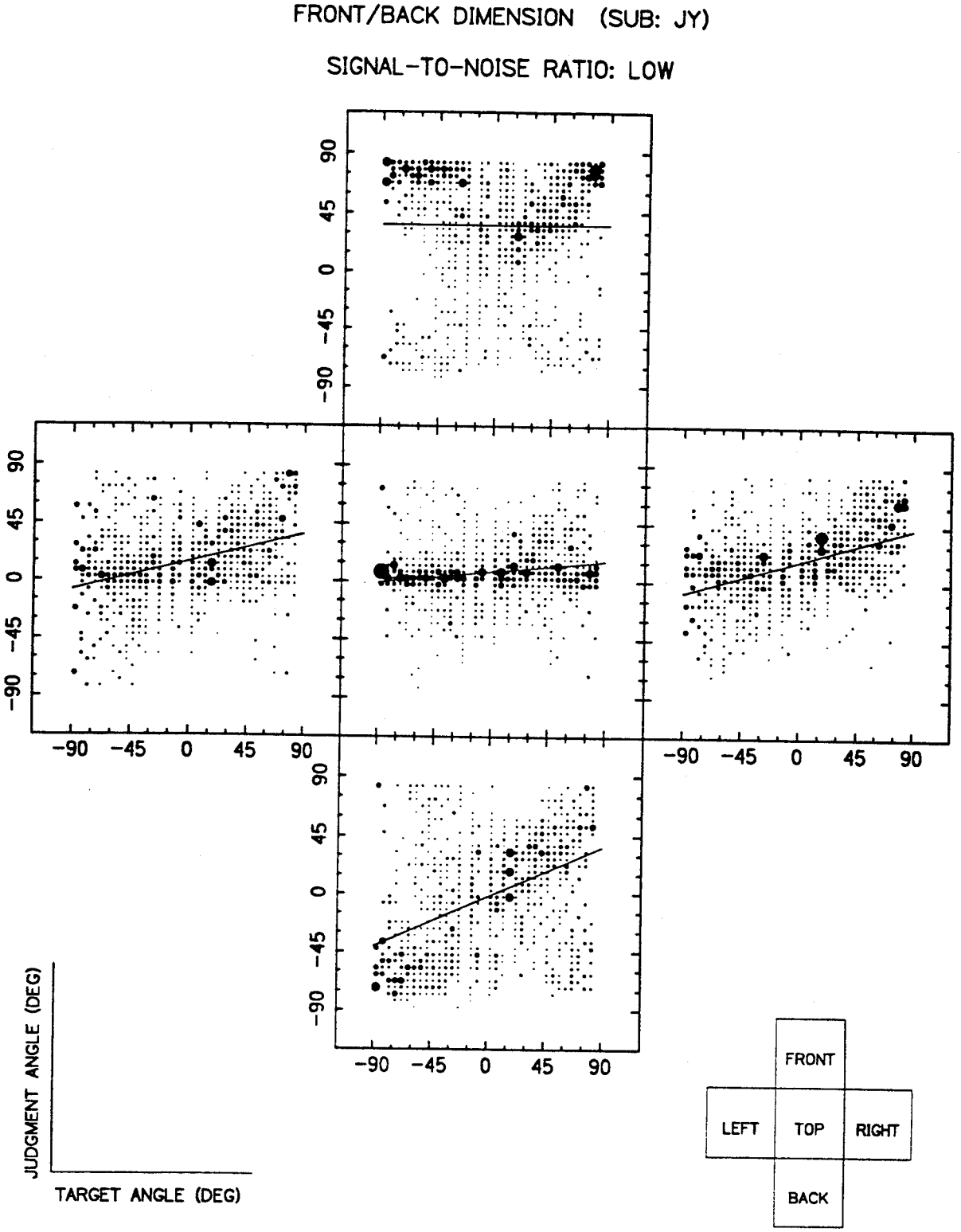
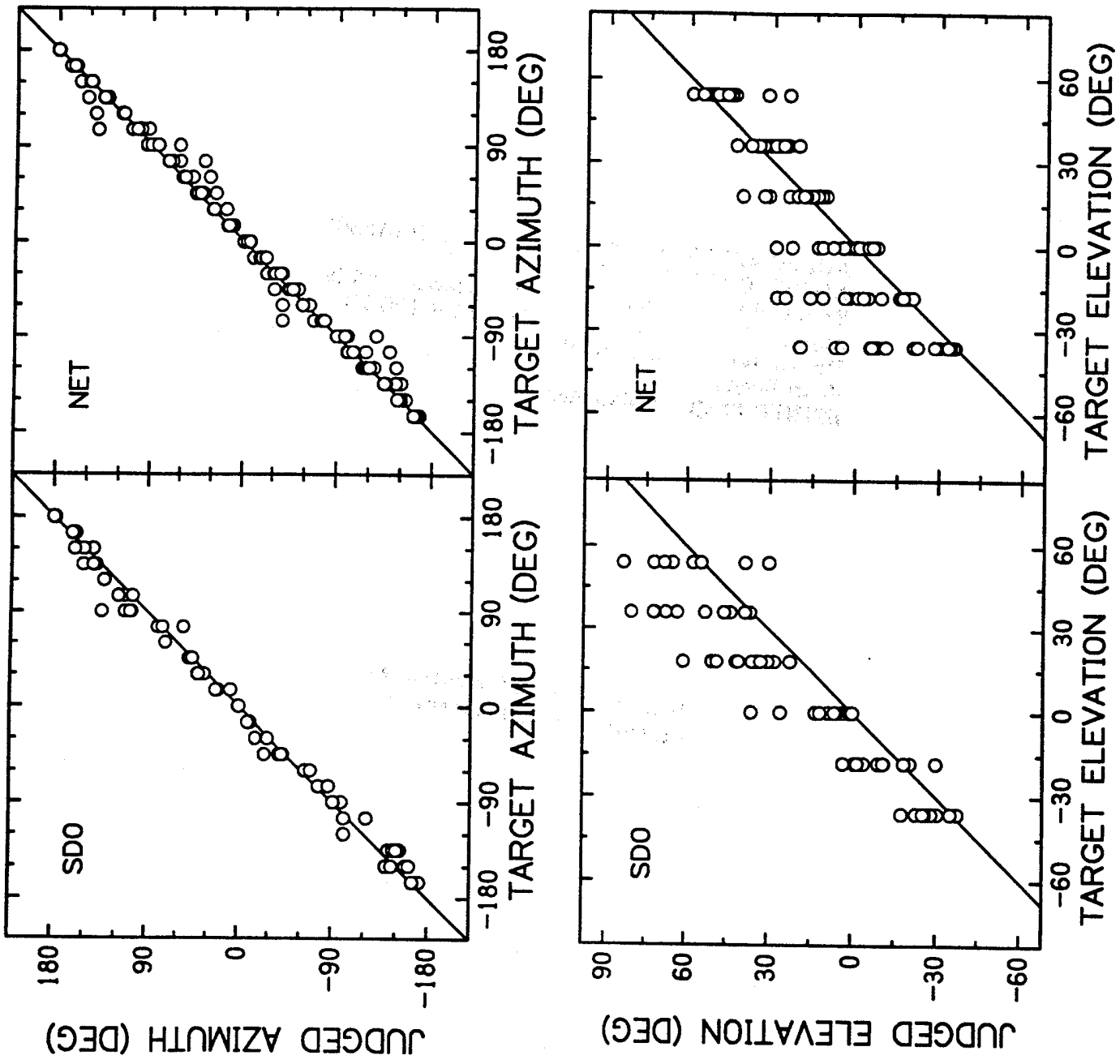

FRONT/BACK DIMENSION (SUB: JY)
SIGNAL-TO-NOISE RATIO: LOW
-90
-45
0
45
90
JUDGMENT ANGLE (DEG)
TARGET ANGLE (DEG)
FRONT
LEFT
TOP
RIGHT
BACK

Figure 9


JUDGED AZIMUTH (DEG)
180
90
0
-90
-180
SDO
NET
-180
-90
0
90
180
TARGET AZIMUTH (DEG)
JUDGED ELEVATION (DEG)
90
60
30
0
-30
-60
SDO
NET
-60
-30
0
30
60
TARGET ELEVATION (DEG)